

# VARIATION ESTIMATES FOR AVERAGES ALONG PRIMES AND POLYNOMIALS

PAVEL ZORIN-KRANICH

**ABSTRACT.** We prove  $q$ -variation estimates,  $q > 2$ , on  $\ell^p$  spaces for averages along primes (with  $1 < p < \infty$ ) and polynomials (with  $|1/p - 1/2| < 1/2(d+1)$ , where  $d$  is the degree of the polynomial). This improves the pointwise ergodic theorems for these averages in the corresponding ranges of  $L^p$  spaces.

## 1. INTRODUCTION

Variation and oscillation estimates for convolution operators associated to polynomials and primes have been pioneered by Bourgain in order to prove the corresponding pointwise ergodic theorems [Bou88c; Bou88a; Bou89]. For the ordinary Cesàro averages the full range of expected variation estimates has been obtained by Jones, Kaufman, Rosenblatt, and Wierdl [JKRW98] and for averages along scalar polynomials variation estimates on  $\ell^2(\mathbb{Z})$  have been obtained by Krause [Kra14a]. Here we prove the corresponding estimates for averages along primes on  $\ell^p(\mathbb{Z})$  and along vector-valued polynomials on  $\ell^p(\mathbb{Z}^d)$  for  $p$  in certain open ranges. See §2 for the definition of the variation norms  $\mathcal{V}^q$  and the relation  $\lesssim$ .

**Theorem 1.1.** *Let*

$$(1.2) \quad K_N = \frac{1}{N} \sum_{n \leq N} \Lambda(n) \delta_n,$$

where  $\Lambda$  denotes the von Mangoldt function. Then for any  $1 < p < \infty$  and  $q > 2$  we have

$$\| \| \| K_N * f \|_{\mathcal{V}_{N \in \mathbb{N}}^q} \|_{\ell^p(\mathbb{Z})} \lesssim_{p,q} \| f \|_{\ell^p(\mathbb{Z})}.$$

**Theorem 1.3.** *Let  $d \geq 1$  and*

$$(1.4) \quad K_N = \frac{1}{N} \sum_{n \leq N} \delta_{(n^1, n^2, \dots, n^d)}.$$

Then for any  $p$  with  $\left| \frac{1}{p} - \frac{1}{2} \right| < \frac{1}{2(d+1)}$  and any  $q > 2$  we have

$$\| \| \| K_N * f \|_{\mathcal{V}_{N \in \mathbb{N}}^q} \|_{\ell^p(\mathbb{Z}^d)} \lesssim_{p,q} \| f \|_{\ell^p(\mathbb{Z}^d)}.$$

Theorem 1.3 is not likely to be optimal as far as the restriction on  $p$  is concerned, and in fact we believe that it should extend to  $1 < p < \infty$ . Interpolation shows that the set of pairs  $(1/p, 1/q)$  for which this result holds is convex, see e.g. [Kra14a, §7]. In view of the maximal inequality (see [MT14, Theorem D] or [Bou89, (7.1)] for the linearly dependent case), corresponding to  $1 < p < \infty$  and  $q = \infty$ , this yields partial results (with a smaller range of  $q$ 's) towards extending the range of allowed  $p$ 's.

The proofs of Theorems 1.1 and 1.3 follow the lines of Bourgain's article [Bou89] but use the more recent variational estimates for convolutions [JSW08] and trigonometric polynomials [NOT10] in order to obtain an appropriate multi-frequency variational inequality on  $L^2(\mathbb{R}^d)$ . Multi-frequency  $L^p$  estimates are obtained in two different ways. The first way consists in interpolation between multi-frequency  $L^2$  estimates and single-frequency

---

2010 *Mathematics Subject Classification.* 42A45 (Primary) 37A45, 26A45 (Secondary).

*Key words and phrases.* variation norm, ergodic averages, prime numbers.

Research supported by the ISF grant 1409/11.

$L^p$  estimates. This approach does not rely on algebraic relations between the distinguished frequencies, but does not yield optimal estimates in our cases. The second way is more specific to our algebraic setting and goes back at least to Wierdl [Wie88], although the lack of an easy endpoint at  $p = \infty$  (as for the maximal inequality) necessitates the use of a more recent transfer technique from [MSW02]. The multi-frequency variational estimates are applied to certain Fourier multipliers that approximate  $\widehat{K_N}$  in a sufficiently strong sense. The construction of these multipliers is due to Bourgain [Bou88b; Bou88a]. We include concise proofs of their properties established in [Bou88a] and [Wie88]. A technical novelty of our argument is that the multi-frequency estimates are used to control the variation norm on a sequence of times which is denser than dyadic. This simplifies the treatment of short variations, see Lemma 4.11.

By Calderón's transference principle [Cal68, Theorem 1] Theorems 1.1 and 1.3 imply the following analogous statements for measure-preserving  $\mathbb{Z}$ -actions on  $\sigma$ -finite measure spaces.

**Corollary 1.5.** *Let  $(X, \mu, T)$  be a  $(\sigma$ -finite, invertible) measure-preserving system. Then for every  $1 < p < \infty$  and  $q > 2$  we have*

$$\left\| \left\| \frac{1}{N} \sum_{n=1}^N \Lambda(n) f(T^n x) \right\|_{\mathcal{V}_{N \in \mathbb{N}}^q} \right\|_{L_x^p} \lesssim_{p,q} \|f\|_{L^p} \quad \text{for every } f \in L^p(X, \mu).$$

**Corollary 1.6.** *Let  $(X, \mu)$  be a  $\sigma$ -finite measure space and  $T_1, \dots, T_d : X \rightarrow X$  be commuting invertible measure-preserving transformations. Then for every  $p$  with  $\left| \frac{1}{p} - \frac{1}{2} \right| < \frac{1}{2(d+1)}$  and every  $q > 2$  we have*

$$\left\| \left\| \frac{1}{N} \sum_{n=1}^N f(T_1^{n^1} \cdots T_d^{n^d} x) \right\|_{\mathcal{V}_{N \in \mathbb{N}}^q} \right\|_{L_x^p} \lesssim_{p,q} \|f\|_{L^p(X)} \quad \text{for every } f \in L^p(X, \mu).$$

Corollary 1.6 applies in particular in the case  $X = \mathbb{Z}$ ,  $T_j = T^{a_j}$ ,  $T : \mathbb{Z} \rightarrow \mathbb{Z}$  the shift. We find it convenient to exclude this non-homogeneous situation from Theorem 1.3 because the homogeneous setup offers a more direct link to the dilation-invariant results from [JSW08].

I thank Mariusz Mirek for pointing out an error in an earlier revision of this text.

## 2. VARIATION OF EXPONENTIAL SUMS

A pointwise variational estimate for an exponential sum with fixed, separated frequencies and varying coefficients was a central technical innovation in the article of Nazarov, Oberlin, and Thiele [NOT10, Lemma 3.2]. We will need a version of this result for exponential sums on  $\mathbb{R}^d$ ,  $d \geq 1$ , with explicit dependence of the constants on all parameters. The higher-dimensional version follows from a more abstract formulation due to Oberlin [Obe13, Proposition 9.3], explicit constants have been given by Krause [Kra14b, Lemma 2.4], and it is clear how these extensions should be combined. We include a detailed proof because our construction of the “parent” function  $\rho$  is slightly simpler than that used in the articles cited above. We begin with a short summary of the relevant definitions.

**Definition 2.1.** Let  $\mathfrak{J}$  be a totally ordered set and  $(c_t)_{t \in \mathfrak{J}}$  be an  $\mathfrak{J}$ -sequence in a normed space. We denote

- (1) by  $\mathcal{N}_\lambda(c)$ ,  $\lambda > 0$ , the *greedy jump counting function*, that is, the supremum over the lengths  $J$  of sequences  $t_0 < t_1 < \cdots < t_J$  such that  $|c_{t_j} - c_{t_{j-1}}| > \lambda$  for all  $j = 1, \dots, J$ ,
- (2) by  $\mathfrak{N}_\lambda(c)$ ,  $\lambda > 0$ , the *lazy jump counting function*, that is, the supremum over the lengths  $J$  of sequences  $s_1 < t_1 \leq s_2 < t_2 \leq \cdots \leq s_J < t_J$  such that  $|c_{t_j} - c_{s_j}| > \lambda$  for all  $j = 1, \dots, J$ ,

- (3) by  $\mathcal{V}^q(c) = \|c_t\|_{\mathcal{V}_t^q}$ ,  $q > 0$ , the *homogeneous  $q$ -variation norm*, that is, the supremum of

$$\|c_{t_{j+1}} - c_{t_j}\|_{\ell_j^q}$$

over all strictly increasing sequences  $t_1 < \dots < t_J$ , and

- (4) the *inhomogeneous  $q$ -variation norm* by

$$\mathcal{V}^q(c) = \|c_t\|_{\mathcal{V}_t^q} = ((\tilde{\mathcal{V}}^q)^q + (\sup_t |c_t|)^q)^{1/q}.$$

We will sometimes write  $\mathcal{N}_{\lambda,t}$ ,  $\mathcal{V}_t^q$ ,  $\mathcal{V}_{t \in \mathcal{J}}^q$ , etc., in order to emphasize the relevant variable and  $\tilde{\mathcal{V}}^q(X)$  in order to emphasize the normed space in which the sequence  $(c_t)$  takes values. It is clear that both  $\mathcal{N}_\lambda$  and  $\mathfrak{N}_\lambda$  are monotonically decreasing in  $\lambda$  and

$$\mathcal{N}_\lambda \leq \mathfrak{N}_\lambda \leq \mathcal{N}_{\lambda/2}.$$

Moreover, we can pass between variation and jump estimates using the identities

$$(2.2) \quad \lambda \mathcal{N}_\lambda^{1/q} \leq \lambda \mathfrak{N}_\lambda^{1/q} \leq \tilde{\mathcal{V}}^q$$

and

$$(2.3) \quad \tilde{\mathcal{V}}^q \leq \left( \sum_{k \in \mathbb{Z}} (2^{k+1})^q \mathfrak{N}_{2^k} \right)^{1/q} \leq 4 \left( \sum_{k \in \mathbb{Z}} (2^k)^q \mathcal{N}_{2^k} \right)^{1/q}.$$

Note that the inhomogeneous variation norm is controlled by the homogeneous variation norm and the value of the sequence at any given point  $t$ . Estimates at a fixed  $t$  will be easy in many of our variation inequalities, allowing us to concentrate on the homogeneous variation norm.

A recurring theme will be splitting the variation into a “long” and a “short” part with respect to an increasing, cofinal, and coinitial sequence  $Z = \{\dots, N_1, N_2, \dots\}$  in  $\mathcal{J}$ . The *long variation* of a sequence  $(c_t)$  with respect to  $Z$  is simply  $\|c_t\|_{\mathcal{V}_{t \in Z}^q}$ . The *short variation* with respect to  $Z$  is defined by

$$\|c_t\|_{\mathcal{S}^q} = \left( \sum_j \|c_t\|_{\mathcal{V}_{t \in [N_j, N_{j+1}]}^q}^q \right)^{1/q}.$$

It is well-known that the full homogeneous variation is controlled by the long and the short variation, namely

$$(2.4) \quad \|c_t\|_{\mathcal{V}_t^q} \leq \|c_t\|_{\mathcal{V}_{t \in Z}^q} + 2\|c_t\|_{\mathcal{S}^q}.$$

To see this, consider any sequence  $t_1 < \dots < t_J$  as in the definition of the homogeneous variation norm. For every  $j$  we have  $t_j \in [N_{j_-}, N_{j_+}]$  with  $j_+ = j_- + 1$ . If  $t_j \leq N_{j_+} \leq N_{(j+1)_-} \leq t_{j+1}$ , then we split the corresponding difference  $c_{t_j} - c_{t_{j+1}}$  accordingly, otherwise we have  $t_{j+1} \in [N_{j_-}, N_{j_+}]$ . Thus the sequence

$$(c_{t_j} - c_{t_{j+1}})_j$$

can be written as the sum of three sequences, one of which corresponds to differences between  $N_j$ 's and the others to differences within intervals  $[N_j, N_{j+1}]$ . Taking the supremum over all increasing sequences of  $t_j$ 's we obtain the claim.

**Lemma 2.5.** *Let  $I = I_1 \times \dots \times I_d \subset \mathbb{R}^d$  be a product of intervals and let  $(\xi_{\vec{k}})_{\vec{k} \in \mathbb{Z}^d} \subset \mathbb{R}^d$  be frequencies such that  $(\xi_{\vec{k},i} - \xi_{\vec{l},i}) \gtrsim |k_i - l_i|/|I_i|$ . Then we have*

$$\left\| \sum_{k \in \mathbb{Z}^d} c_k e(\xi_k \cdot y) \right\|_{L_y^2(I)} \lesssim |I|^{1/2} \|c_k\|_{\ell_k^2},$$

where the implied constant depends only on the implied constant in the hypothesis and the dimension  $d$ .

Here and later  $C_a$  denotes an unspecified positive constant, depending on auxiliary parameter(s)  $a$ , whose value may vary from line to line. We say that  $A$  is dominated by  $B$ , in symbols  $A \lesssim_a B$ , if  $A \leq C_a B$ . The parameters  $a$  can be partially or fully omitted if they are clear from the context.

*Proof.* Let  $w_i$  be a smooth non-negative functions bounded by 1 and supported on  $2I_i$  with  $w_i|_{I_i} \equiv 1$  and  $|w_i''||I_i|^2 \lesssim 1$ . Let also  $w(y) = w_1(y_1) \cdots w_d(y_d)$ . We use almost-orthogonality of the phases  $e(\xi_k \cdot y)$  in  $L^2(w)$ . More precisely, by partial integration we obtain

$$\begin{aligned} \left\| \sum_k c_k e(\xi_k \cdot y) \right\|_{L_y^2(I)}^2 &\leq \left\| \sum_k c_k e(\xi_k \cdot y) \right\|_{L_y^2(w)}^2 \leq \sum_{k,l \in \mathbb{Z}^d} \left| c_k \bar{c}_l \int e((\xi_k - \xi_l) \cdot y) w(y) dy \right| \\ &\leq \sum_{k,l \in \mathbb{Z}^d} |c_k c_l| \left| \prod_{i: k_i \neq l_i} ((2\pi i)(\xi_{k,i} - \xi_{l,i}))^{-2} \int e((\xi_j - \xi_k) \cdot y) \left( \prod_{i: k_i \neq l_i} \partial_i^2 \right) w(y) dy \right| \\ &\lesssim \sum_{k,l \in \mathbb{Z}^d} |c_k c_l| \left( \prod_{i: k_i \neq l_i} |I_i|^2 (k_i - l_i)^{-2} \right) |I| \left( \prod_{i: k_i \neq l_i} |I_i|^{-2} \right) \\ &\leq |I| \sum_{k,l \in \mathbb{Z}^d} (|c_k|^2 + |c_l|^2) / 2 \prod_{i: k_i \neq l_i} (k_i - l_i)^{-2} \end{aligned}$$

Since the last expression is symmetric in  $k$  and  $l$ , it is bounded by

$$|I| \sum_{k,l \in \mathbb{Z}^d} |c_k|^2 \prod_{i: k_i \neq l_i} (k_i - l_i)^{-2} = |I| \sum_{k \in \mathbb{Z}^d} |c_k|^2 \sum_{l \in \mathbb{Z}^d} \prod_{i: k_i \neq l_i} (k_i - l_i)^{-2} = |I| \sum_{k \in \mathbb{Z}^d} |c_k|^2 \sum_{l \in \mathbb{Z}^d} \prod_{i: l_i \neq 0} l_i^{-2}.$$

Since the last sum over  $l$  is finite, we obtain the claim.  $\square$

The next lemma captures the main step in the proof of [NOT10, Lemma 3.2]. In the formulation below the left-hand side is essentially from [Obe13, Proposition 9.3] and the right-hand side is essentially from [Kra14b, Lemma 2.4].

**Lemma 2.6.** *Let  $B$  be a normed space,  $I$  a measure space, and let  $g \in L^r(I, B')$ ,  $r \geq 1$ . Let also  $(c_t)_{t \in \mathfrak{J}} \subset B$  with a countable totally ordered set  $\mathfrak{J}$ , and  $q > r$ . Then*

$$\| \| \langle c_t, g(y) \rangle \|_{\mathcal{N}_t^q} \|_{L_y^r(I)} \lesssim \int_0^\infty \min(M \mathcal{N}_\lambda^{1/r}, \|g\|_{L^r(I, B')} \mathcal{N}_\lambda^{1/q}) d\lambda,$$

where  $\mathcal{N}$  is the greedy jump counting function for the sequence  $(c_t)$ ,

$$M := \sup_{c \in B, \|c\|=1} \| \langle c, g(y) \rangle \|_{L_y^r(I)},$$

and the implied constant is absolute.

*Proof.* It suffices to consider finite sequences  $(c_t)_{t=1}^T$  as long as the bounds do not depend on  $T$ . We may assume that the minimal jump size  $\min_{t < T} \|c_t - c_{t+1}\|_B > \lambda > 0$ , otherwise one can remove some of the terms from the sequence  $(c_t)$ . We construct a sequence of increasingly coarse partitions of  $\{1, \dots, T\}$  into blocks with bounded  $\infty$ -variation and jumps between blocks in such a way that both the upper bounds on the  $\infty$ -variation and the lower bounds on the jumps increase exponentially. To this end we recursively define a sequence of functions  $\rho(n, \cdot) : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ . We begin with

$$\rho(0, t) = t.$$

Suppose that  $\rho(n, \cdot)$  has been defined for some  $n$  and define  $\rho(n+1, t)$  by recursion in  $t$  starting with  $\rho(n+1, 1) = 1$  by

$$\rho(n+1, t+1) := \begin{cases} \rho(n+1, t) & \text{if } \|c_{\rho(n+1, t)} - c_{\rho(n, t+1)}\|_B \leq 2^{n+1}\lambda, \\ \rho(n, t+1) & \text{otherwise.} \end{cases}$$

It follows that  $\rho(n, t)$  is monotonically increasing in  $t$  and monotonically decreasing in  $n$ . Moreover, for all  $n$  and  $t$  we have

$$(2.7) \quad \|c_{\rho(n, t)} - c_{\rho(n, t+1)}\|_B > 2^n \lambda \quad \text{provided } \rho(n, t) \neq \rho(n, t+1),$$

$$(2.8) \quad \|c_{\rho(n, t)} - c_{\rho(n+1, t)}\|_B \leq 2^{n+1} \lambda,$$

$$(2.9) \quad \rho(n+1, t+1) \neq \rho(n+1, t) \implies \rho(n, t+1) \neq \rho(n, t).$$

The implication (2.9) can be easily seen by the contrapositive and a case distinction in the definition of  $\rho(n+1, t)$ . Note that (2.7) implies  $\rho(n, t) = 1$  for all  $t$  if  $n$  is sufficiently large. Write

$$c_t = c_1 + \sum_{n=0}^{\infty} (c_{\rho(n, t)} - c_{\rho(n+1, t)}).$$

By subadditivity of the homogeneous variation norm we have

$$\| \| \langle c_t, g(y) \rangle \|_{\tilde{\mathcal{V}}_t^q} \|_{L_y^r(I)} \leq \sum_{n=0}^{\infty} \| \| \langle c_{\rho(n, t)} - c_{\rho(n+1, t)}, g(y) \rangle \|_{\tilde{\mathcal{V}}_t^q} \|_{L_y^r(I)}.$$

For each  $n$  we estimate the corresponding summand. Observe that the lower bound on the jump size in (2.7) implies that the sequence  $\rho(n, \cdot)$  makes at most  $\mathcal{N}_{2^n \lambda}$  jumps, before places  $J_n \subset \{1, \dots, T\}$ , say. Note that  $J_{n+1} \subset J_n$  by (2.9). Hence the variation norm in the summand collapses to the subsequence

$$J'_n = \begin{cases} \{1\} \cup J_n & \text{if } J_n \neq \emptyset, \\ \emptyset & \text{otherwise.} \end{cases}$$

On that subsequence we estimate the  $\tilde{\mathcal{V}}^q$  norm by the  $\ell^q$  norm, thereby obtaining the following bound for the  $n$ -th summand:

$$(2.10) \quad \| \| \langle c_{\rho(n, t)} - c_{\rho(n+1, t)}, g(y) \rangle \|_{\ell_{t \in J'_n}^q} \|_{L_y^r(I)}.$$

The first way to proceed from here is to estimate the  $\ell^q$  norm by the  $\ell^r$  norm and to change the order of integration (in  $t$  and  $y$ ). Using (2.8) this gives the bound

$$M 2^{n+1} \lambda |J'_n|^{1/r} \lesssim M 2^n \lambda \mathcal{N}_{2^n \lambda}^{1/r}.$$

The second way to proceed is to estimate the dual pairing by the product of norms, which gives for (2.10) the estimate

$$\| \| \| c_{\rho(n, t)} - c_{\rho(n+1, t)} \|_B \| g(y) \|_{B'} \|_{\ell_{t \in J'_n}^q} \|_{L_y^r(I)}$$

By (2.8) this gives the bound

$$\| g \|_{L^r(I, B')} 2^{n+1} \lambda |J'_n|^{1/q} \lesssim \| g \|_{L^r(I, B')} 2^n \lambda \mathcal{N}_{2^n \lambda}^{1/q}.$$

Combining these estimates we obtain

$$\| \| \langle c_t, g(y) \rangle \|_{\tilde{\mathcal{V}}_t^q} \|_{L_y^r(I)} \lesssim \sum_{n \in \mathbb{N}} 2^n \lambda \min(M \mathcal{N}_{2^n \lambda}^{1/r}, \| g \|_{L^r(I, B')} \mathcal{N}_{2^n \lambda}^{1/q}),$$

and the claim follows by monotonicity of the jump counting function.  $\square$

**Corollary 2.11** (cf. [NOT10, Lemma 3.2]). *Let  $G \subset \mathbb{Z}^d$  be a set of size  $N$  and  $(\xi_l)_{l \in G} \subset \mathbb{R}^d$ ,  $I_i$ ,  $I$  be as in Lemma 2.5. For any  $2 < r < q$  and any sequence  $(c_t) \subset \ell^2[G]$  we have*

$$\| \sum_{l \in G} c_{t,l} e(\xi_l \cdot y) \|_{\tilde{\mathcal{V}}_t^q} \|_{L_y^2(I)} \lesssim \left( \frac{q}{q-r} + \frac{2}{r-2} \right) |I|^{1/2} N^{(\frac{1}{2} - \frac{1}{r}) \frac{q}{q-2}} \|c_t\|_{\tilde{\mathcal{V}}_t^r(\ell^2[G])},$$

where the implied constant is absolute.

*Proof.* We apply Lemma 2.6 with  $B = B' = \ell^2[G]$ ,  $r = 2$ , and  $g(y) = (e(\xi_l \cdot y))_{l \in G}$ . Then  $\|g\|_{L^2(I, B')} = N^{1/2} |I|^{1/2}$  and  $M \lesssim |I|^{1/2}$  by Lemma 2.5. Hence we obtain

$$\| \sum_{l \in G} c_{t,l} e(\xi_l \cdot y) \|_{\tilde{\mathcal{V}}_t^q} \|_{L_y^2(I)} \lesssim |I|^{1/2} \int_0^\infty \min(\mathcal{N}_\lambda^{1/2}, N^{1/2} \mathcal{N}_\lambda^{1/q}) d\lambda.$$

We have  $\mathcal{N}_\lambda(c) \leq a^r / \lambda^r$  with  $a = \|c_t\|_{\tilde{\mathcal{V}}_t^r(\ell^2[G])}$ . Splitting the integral at  $\lambda_0 = aN^{-1/(2r(1/2-1/q))}$  we obtain

$$\begin{aligned} & \int_0^{\lambda_0} N^{1/2} (a^r / \lambda^r)^{1/q} d\lambda + \int_{\lambda_0}^\infty (a^r / \lambda^r)^{1/2} d\lambda \\ &= N^{1/2} a^{r/q} (-r/q + 1)^{-1} \lambda_0^{-r/q+1} - a^{r/2} (-r/2 + 1)^{-1} \lambda_0^{-r/2+1} \\ &= aN^{(\frac{1}{2} - \frac{1}{r}) \frac{q}{q-2}} ((1 - r/q)^{-1} + (r/2 - 1)^{-1}). \quad \square \end{aligned}$$

### 3. FOURIER MULTIPLIERS ON $\mathbb{R}^d$

The main result of this section, Proposition 3.9, is a multiple-frequency variation inequality on  $L^2(\mathbb{R}^d)$  with a good (logarithmic) dependence of the bounds on the number of frequencies involved in it. We begin by recalling several variation inequalities due to Jones, Seeger, and Wright, limiting ourselves to the minimal level of generality required in our applications. The first is a special case of [JSW08, Lemma 2.1], which goes back to Bourgain's argument from [Bou89, §3].

**Lemma 3.1.** *Let  $(X, \mu)$  be a measure space and  $(T_i)_{i \in \mathbb{J} \subset \mathbb{R}}$  be a family of continuous linear operators on  $L^p(X)$ ,  $1 < p < \infty$ , that are contractive on  $L^\infty(X)$  and such that  $T_i f(x)$  is continuous in  $i$  for almost every  $x$ . Suppose that*

$$(3.2) \quad \sup_{\lambda > 0} \|\lambda(\mathcal{N}_{\lambda, i}(T_i f)(x))^{1/2}\|_{L^r(X, \mu)} \lesssim_r \|f\|_{L^r(X, \mu)}$$

for every  $1 < r < \infty$  and every characteristic function  $f = \chi_A$  of a finite measure subset  $A \subset X$ . Then for every  $q > 2$  and  $1 < p < \infty$  we have

$$(3.3) \quad \| \|T_i f(x)\|_{\tilde{\mathcal{V}}_{i \in \mathbb{J}}^q} \|_{L^p(X, \mu)} \lesssim_p \frac{q}{q-2} \|f\|_{L^p(X, \mu)}.$$

Let us point out how the various qualitative assumptions are used in the proof of [JSW08, Lemma 2.1]. By the qualitative assumption of pointwise continuity almost everywhere the problem reduces to countable index sets  $I \subset \mathbb{R}$ , and in particular the jump counting functions and the pointwise variation norms become measurable functions on  $X$ . This in turn allows one to use monotone convergence to reduce the problem to finite sets  $I$ . The proof proceeds by establishing restricted strong type estimates, which are then interpolated to the requested strong type estimates. However, these are a priori obtained for simple functions (finite linear combinations of characteristic functions), and the qualitative assumption of  $L^p$  continuity of the individual operators is needed to pass to the full  $L^p$  space.

**3.1. A variation inequality for a single frequency.** We will apply Lemma 3.1 in the setting of convolution operators. Let  $\sigma_t$  be the measure on  $\mathbb{R}^d$  defined by

$$(3.4) \quad \int f d\sigma_t = \frac{1}{t} \int_{s=0}^t f(s^1, s^2, \dots, s^d) ds.$$

The following result is stated in a remark following [JSW08, Theorem 1.5].

**Theorem 3.5.** *For any  $1 < p < \infty$  we have*

$$\sup_{\lambda > 0} \|\lambda \mathfrak{N}_{\lambda, t \in \mathbb{R}}(f * \sigma_t(x))^{1/2}\|_{L_x^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}.$$

**Corollary 3.6.** *For any  $1 < p < \infty$  and any  $s > 2$  we have*

$$\| \|f * \sigma_t(x)\|_{\mathcal{V}_{t>0}^s} \|_{L_x^p(\mathbb{R}^d)} \lesssim_p \frac{s}{s-2} \|f\|_{L^p(\mathbb{R}^d)},$$

where the implied constant does not depend on  $s$ .

*Proof.* One can show that the function  $s \mapsto f(x_1 + s^1, \dots, x_d + s^d)$  is locally integrable for almost every  $(x_1, \dots, x_d)$  using Fubini's theorem. This implies that  $t \mapsto f * \sigma_t(x)$  is continuous in  $t$ , and in view of Theorem 3.5 we may apply Lemma 3.1.  $\square$

This implies the following variation version of [Bou89, Lemma 3.30].

**Corollary 3.7.** *For any  $2 \leq p < \infty$ , any set  $G$ , and any  $s > 2$  we have*

$$\| \|f_l * \sigma_t(x)\|_{\mathcal{V}_{t>0}^s(\ell_{l \in G}^2)} \|_{L_x^p(\mathbb{R}^d)} \lesssim_p \frac{s}{s-2} \| \|f_l\|_{L^p(\mathbb{R}^d)} \|_{\ell_{l \in G}^2},$$

where the implied constant does not depend on  $G$  and  $s$ .

*Proof.* The case  $|G| = 1$  is given by Corollary 3.6.

In the general case for any finite sequence  $t_1 < \dots < t_J$  and  $x \in \mathbb{R}$  we have

$$\begin{aligned} \|f_l * \sigma_{t_j}(x) - f_l * \sigma_{t_{j-1}}(x)\|_{\ell_j^s(\ell_l^2)} &= \| \|f_l * \sigma_{t_j}(x) - f_l * \sigma_{t_{j-1}}(x)\|_{\ell_l^2} \|_{\ell_j^s} \\ &\leq \| \|f_l * \sigma_{t_j}(x) - f_l * \sigma_{t_{j-1}}(x)\|_{\ell_j^s} \|_{\ell_l^2} \leq \| \|f_l * \sigma_t(x)\|_{\mathcal{V}_t^s} \|_{\ell_l^2} \end{aligned}$$

by the Minkowski inequality and the assumption  $s > 2$ . Taking the supremum over all increasing finite sequences we obtain

$$(3.8) \quad \|f_l * \sigma_t(x)\|_{\mathcal{V}_t^s(\ell_l^2)} \leq \| \|f_l * \sigma_t(x)\|_{\mathcal{V}_t^s} \|_{\ell_l^2}.$$

Integrating this we obtain

$$\| \|f_l * \sigma_t(x)\|_{\mathcal{V}_t^s(\ell_l^2)} \|_{L_x^p} \leq \| \| \|f_l * \sigma_t(x)\|_{\mathcal{V}_t^s} \|_{\ell_l^2} \|_{L_x^p}.$$

By the Minkowski inequality and the assumption  $p \geq 2$  this is bounded by

$$\| \| \|f_l * \sigma_t(x)\|_{\mathcal{V}_t^s} \|_{L_x^p} \|_{\ell_l^2}.$$

Using the case  $|G| = 1$  we obtain the conclusion.  $\square$

**3.2. A variation inequality for several frequencies.** A central observation is that Corollaries 2.11 and 3.7 can be used to show a multi-frequency variation inequality in the same manner as in [Bou89, Lemma 4.13].

**Proposition 3.9.** *Let  $G \subset \mathbb{Z}^d$ ,  $|G| = N$ , and  $(\xi_l)_{l \in G} \subset \mathbb{R}^d$  be frequencies such that  $|\xi_{l,i} - \xi_{l',i}| > |l_i - l'_i| \tau_i$ ,  $\tau_i > 0$ , for all  $i = 1, \dots, d$  and  $l, l' \in G$ . Let  $f_l \in L^2(\mathbb{R}^d)$ ,  $l \in G$ , be functions with  $\text{supp } \hat{f}_l \subset [-\tau_1/2, \tau_1/2] \times \dots \times [-\tau_d/2, \tau_d/2]$ . Then for any  $q > 2$  we have*

$$(3.10) \quad \| \sum_{l \in G} e(\xi_l \cdot x) (f_l * \sigma_t)(x) \|_{\mathcal{V}_{t>0}^q} \|_{L_x^2(\mathbb{R}^d)} \lesssim \left( \frac{q(\log N + 1)}{q - 2} \right)^2 \| \|f_l\|_{L^2(\mathbb{R}^d)} \|_{\ell_{l \in G}^2}.$$

*Proof.* When the variation norm on the left-hand side of (3.10) is replaced by evaluation at  $t = 1$ , say, the  $L^2$  bound follows from the Plancherel identity. Hence it suffices to show (3.10) with the homogeneous variation norm  $\tilde{\mathcal{V}}^q$ .

It suffices to consider  $N > 100$  and  $q < 4$ , say. As in the proof of Corollary 3.7 we may restrict  $t$  in (3.10) to the rationals, and by monotone convergence it suffices to consider a finite subset  $T$  of the rationals as long as the bounds are independent of this set.

Let  $B_T$  be the best constant for which the restricted version of (3.10) on  $t \in T$  holds. It is finite because we can estimate the  $\tilde{\mathcal{V}}^q$  norm by the  $\ell^2$  norm, thereby bounding the left-hand side of (3.10) by

$$\sum_{l \in G} |||f_l * \sigma_t||_{\ell_{t \in T}^2}^2 |||f_l||_{L^2} \leq \sum_{l \in G} |||f_l * \sigma_t||_{L^2}^2 |||f_l||_{\ell_{t \in T}^2}^2 \leq C \sum_{l \in G} |T|^{1/2} |||f_l||_{L^2} \leq CN^{1/2} |T|^{1/2} |||f_l||_{L^2} |||f_l||_{\ell_{l \in G}^2}^2$$

using the Minkowski, the Young convolution, and the Hölder inequality. We now use Bourgain's averaging trick. Let  $R_u f(x) = f(x + u)$ . By the frequency support assumption on  $f_l$  and the Bernstein inequality we have  $||f_l - R_u f_l||_{L^2} \lesssim \sum_i \tau_i |u_i| ||f_l||_{L^2}$ . It follows that

$$\begin{aligned} & ||| \sum_{l \in G} e(\xi_l \cdot x) (f_l * \sigma_t)(x) |||_{\tilde{\mathcal{V}}_{t \in T}^q}^2 |||f_l||_{L_x^2}^2 \\ & \leq |I|^{-1} ||| \int_I \sum_{l \in G} e(\xi_l \cdot x) (R_u f_l * \sigma_t)(x) du |||_{\tilde{\mathcal{V}}_{t \in T}^q}^2 |||f_l||_{L_x^2}^2 + \frac{B_T}{2} |||f_l||_{L^2}^2 |||f_l||_{\ell_{l \in G}^2}^2, \end{aligned}$$

where  $I = I_1 \times \cdots \times I_d$ ,  $I_i = [-c/\tau_i, c/\tau_i]$ , and  $c$  is a sufficiently small constant depending only on  $d$ . Pulling the integral out of the variation norm and estimating the  $L_u^1$  norm by the  $L_u^2$  norm we obtain the bound

$$C |I|^{-1/2} ||| \sum_{l \in G} e(\xi_l \cdot x) (R_u f_l * \sigma_t)(x) |||_{\tilde{\mathcal{V}}_{t \in T}^q}^2 |||f_l||_{L_u^2(I)}^2 |||f_l||_{L_x^2}^2$$

for the first summand. The fact that the translation operator  $R$  commutes with convolution and a change of variable in the double integral in  $x$  and  $u$  show that this equals

$$C |I|^{-1/2} ||| \sum_{l \in G} e(\xi_l \cdot (x + u)) (f_l * \sigma_t)(x) |||_{\tilde{\mathcal{V}}_{t \in T}^q}^2 |||f_l||_{L_u^2(I)}^2 |||f_l||_{L_x^2}^2.$$

By Corollary 2.11 for any  $2 < r < q$  this is bounded by

$$\begin{aligned} & C \left( \frac{q}{q-r} + \frac{2}{r-2} \right) N^{(\frac{1}{2}-\frac{1}{r})\frac{q}{q-2}} |||e(\xi_l \cdot x) (f_l * \sigma_t)(x) |||_{\tilde{\mathcal{V}}_{t \in T}^r(\ell_{l \in G}^2)}^2 |||f_l||_{L_x^2}^2 \\ & = C \left( \frac{q}{q-r} + \frac{2}{r-2} \right) N^{(\frac{1}{2}-\frac{1}{r})\frac{q}{q-2}} |||(f_l * \sigma_t)(x) |||_{\tilde{\mathcal{V}}_{t \in T}^r(\ell_{l \in G}^2)}^2 |||f_l||_{L_x^2}^2. \end{aligned}$$

By Corollary 3.7 this is bounded by

$$C \left( \frac{q}{q-r} + \frac{2}{r-2} \right) N^{(\frac{1}{2}-\frac{1}{r})\frac{q}{q-2}} \frac{r}{r-2} |||f_l||_{L^2}^2 |||f_l||_{\ell_{l \in G}^2}^2.$$

Choosing  $r$  such that  $r - 2 = (q - 2)(\log N)^{-1}$  this gives the bound

$$C \left( \frac{q \log N}{q - 2} \right)^2 |||f_l||_{L^2}^2 |||f_l||_{\ell_{l \in G}^2}^2.$$

Hence we have obtained

$$B_T \leq C \left( \frac{q \log N}{q - 2} \right)^2 + \frac{B_T}{2},$$

and the conclusion follows.  $\square$



**3.3.  $L^p$  variation estimates.** Let us now describe the setting in which Proposition 3.9 will be applied. Let  $\eta$  be a Schwartz cut-off function such that  $\text{supp } \widehat{\eta} \subset [-1/50, 1/50]^d$  and  $\widehat{\eta} \equiv 1$  on  $[-1/100, 1/100]^d$ . Let

$$\mathcal{R}_s = \{(a_1/q_1, \dots, a_d/q_d) \in \mathbb{Q}^d \cap [0, 1]^d \text{ in reduced form with } 2^s \leq \text{lcm}(q_1, \dots, q_d) < 2^{s+1}\}$$

be the set of rational points of height  $\approx 2^s$  in the unit cube.

Let  $L_{s,t}^{\mathbb{R}}$  be Schwartz functions on  $\mathbb{R}^d$  defined by

$$(3.11) \quad \widehat{L_{s,t}^{\mathbb{R}}}(\vec{\alpha}) = \sum_{\vec{\theta} \in \mathcal{R}_s} S(\vec{\theta}) \widehat{\sigma_t}(\vec{\alpha} - \vec{\theta}) \widehat{\eta_{10^s}}(\vec{\alpha} - \vec{\theta}),$$

where  $S(\vec{\theta})$  are arbitrary constants and  $\eta_t(\vec{x}) = t^{-d} \eta(\vec{x}/t)$  denotes the  $L^1$ -dilation, so that  $\widehat{\eta_t}(\vec{\xi}) = \widehat{\eta}(t\vec{\xi})$ . An important observation is that any two distinct members of  $\mathcal{R}_s$  are separated at least by  $2^{-2s-2}$ , so the terms of the sum defining  $\widehat{L_{s,t}^{\mathbb{R}}}$  are disjointly supported. Write

$$(3.12) \quad \mathbf{S}_s = \max_{\vec{\theta} \in \mathcal{R}_s} S(\vec{\theta}).$$

In our applications this quantity will decrease with  $s$  sufficiently rapidly to offset the relatively fast growth of the size of the set of frequencies  $\mathcal{R}_s$ , thus making the next result useful at least for  $p$  not too far from 2.

**Theorem 3.13.** *For any  $1 < p < \infty$ , any  $q > 2$ , and any  $\delta > 0$  we have*

$$(3.14) \quad |||(L_{s,t}^{\mathbb{R}} * f)(x)| |_{\gamma_{t>0}^q} ||_{L_x^p(\mathbb{R}^d)} \lesssim_{q,p,\delta} \mathbf{S}_s |\mathcal{R}_s|^{2|\frac{1}{p}-\frac{1}{2}|+\delta} \|f\|_{L^p(\mathbb{R}^d)}.$$

*Proof.* We have

$$\begin{aligned} (L_{s,t}^{\mathbb{R}} * f)(x) &= \sum_{\vec{\theta} \in \mathcal{R}_s} S(\vec{\theta}) \mathcal{F}^{-1}(\widehat{\sigma_t}(\cdot - \vec{\theta}) \widehat{\eta_{10^s}}(\cdot - \vec{\theta}) \widehat{f})(x) \\ &= \sum_{\vec{\theta} \in \mathcal{R}_s} S(\vec{\theta}) e(\vec{\theta} \cdot x) \mathcal{F}^{-1}(\widehat{\sigma_t} \widehat{\eta_{10^s}} \widehat{f}(\cdot + \vec{\theta}))(x) \\ &= \sum_{\vec{\theta} \in \mathcal{R}_s} e(\vec{\theta} \cdot x) (\sigma_t * f_{\vec{\theta}})(x) \end{aligned}$$

with  $f_{\vec{\theta}} = S(\vec{\theta}) \eta_{10^s} * (e(-\vec{\theta} \cdot \cdot) f)$ . By the Plancherel identity we have

$$(3.15) \quad |||f_{\vec{\theta}}||_{L^2} ||_{\ell_{\vec{\theta}}^2} \lesssim \mathbf{S}_s \|f\|_{L^2},$$

whereas by the Young convolution inequality we have

$$(3.16) \quad \|f_{\vec{\theta}}\|_{L^p} \lesssim |S(\vec{\theta})| \|f\|_{L^p} \quad \text{for any } 1 \leq p \leq \infty.$$

Using (3.15) and Proposition 3.9 we obtain

$$(3.17) \quad |||(L_{s,t}^{\mathbb{R}} * f)(x)| |_{\gamma_{t>0}^q} ||_{L_x^2(\mathbb{R}^d)} \lesssim \mathbf{S}_s \left( \frac{q(\log |\mathcal{R}_s| + 1)}{q - 2} \right)^2 \|f\|_{L^2(\mathbb{R}^d)} \lesssim_{q,\delta} \mathbf{S}_s |\mathcal{R}_s|^{\delta} \|f\|_{L^2(\mathbb{R}^d)}$$

for any  $\delta > 0$ , where the implied constant does not depend on  $s$ . This is the conclusion for  $p = 2$ .

In view of (3.17) and by interpolation it suffices to establish the conclusion with  $|\mathcal{R}_s|^{2|\frac{1}{p}-\frac{1}{2}|+\delta}$  replaced by  $|\mathcal{R}_s|$ . This follows from the single-frequency estimate

$$|||\sigma_t * f_{\vec{\theta}}||_{\gamma_{t>0}^q} ||_{L^p(\mathbb{R}^d)} \lesssim_{q,p} \|f_{\vec{\theta}}\|_{L^p}$$

given by Corollary 3.6 and the easy estimate at  $t = 1$ . □

4. FOURIER MULTIPLIERS ON  $\mathbb{Z}^d$ 

**4.1. Transfer from the reals.** Estimates for Fourier multipliers on the real line can be transferred to the integers by a standard averaging argument. A particularly useful version of that argument, due to Magyar, Stein, and Wainger, shows that the loss in the operator norm is uniformly bounded over all  $L^p$  spaces and, for operator-valued multipliers, over the Banach spaces in the fibers.

**Theorem 4.1** ([MSW02, Corollary 2.1]). *Let  $B_1, B_2$  be finite-dimensional Banach spaces and  $m : \mathbb{R}^d \rightarrow L(B_1, B_2)$  a bounded function supported on a cube with side length 1 containing the origin that acts as a Fourier multiplier from  $L^p(\mathbb{R}^d, B_1)$  to  $L^p(\mathbb{R}^d, B_2)$  for some  $1 \leq p \leq \infty$ . Let  $q$  be a positive integer and*

$$m_{\text{per}}^q(\xi) := \sum_{l \in \mathbb{Z}^d} m(q\xi - l) \quad \text{for } \xi \in (\mathbb{R}/\mathbb{Z})^d.$$

*Then  $m_{\text{per}}^q$  acts as a Fourier multiplier from  $\ell^p(\mathbb{Z}^d, B_1)$  to  $\ell^p(\mathbb{Z}^d, B_2)$  with norm*

$$\|m_{\text{per}}^q\|_{\ell^p(\mathbb{Z}^d, B_1) \rightarrow \ell^p(\mathbb{Z}^d, B_2)} \lesssim_d \|m\|_{L^p(\mathbb{R}^d, B_1) \rightarrow L^p(\mathbb{R}^d, B_2)}.$$

*The implied constant does not depend on  $p, q, B_1$ , and  $B_2$ .*

This allows us to transfer Theorem 3.13 to the following statement on sequence spaces.

**Proposition 4.2.** *Let  $1 < p < \infty$ ,  $q > 2$ , and  $\delta > 0$  be arbitrary. Then*

$$(4.3) \quad \|\widehat{L_{s,t} * f}\|_{\mathcal{V}_{t>0}^q} \|f\|_{\ell^p(\mathbb{Z}^d)} \lesssim_{p,q,\delta} |\mathcal{R}_s|^{2|\frac{1}{p}-\frac{1}{2}|+\delta} \|f\|_{\ell^p(\mathbb{Z}^d)},$$

*where  $\widehat{L_{s,t}}$  is defined by (3.11) as a function on  $(\mathbb{R}/\mathbb{Z})^d$ .*

More precisely, note that the convolution on the left-hand side is, pointwise, a continuous function of  $t$ , so we may restrict attention to rational  $t$ . By monotone convergence it suffices to consider a finite set  $T$  of  $t$ 's as long as we obtain an estimate that does not depend on that set. We apply Theorem 4.1 with  $B_1 = \mathbb{C}$ ,  $B_2 = (\mathbb{C}^T, \mathcal{V}_{t \in T}^q)$ , and  $q = 1$ .

**4.2. Long variation.** In this section we estimate the long variation for convolutions with kernels that admit favorable approximation in terms of

$$(4.4) \quad L_N = \sum_{s \geq 0} L_{s,N}.$$

**Theorem 4.5.** *Let  $d \geq 1$ ,  $1 < p < \infty$ ,  $Z \subset \mathbb{N}$ , and suppose*

$$(4.6) \quad \sum_{N \in Z} \|\widehat{K_N} - \widehat{L_N}\|_{\infty}^{\delta} < \infty \quad \text{for every } \delta > 0 \quad \text{and}$$

$$(4.7) \quad \sum_{s \geq 0} |\mathcal{R}_s|^{2|\frac{1}{p}-\frac{1}{2}|+\delta} \mathbf{s}_s < \infty \quad \text{for some } \delta > 0.$$

*If  $p \neq 2$  assume in addition*

$$(4.8) \quad \sup_{N \in Z} \|K_N\|_{\ell^1} < \infty.$$

*Then for every  $q > 2$  we have*

$$\|\widehat{K_N} * f\|_{\mathcal{V}_{N \in Z}^q} \|f\|_{\ell^p} \lesssim_q \|f\|_{\ell^p}.$$

*Proof.* We have

$$(4.9) \quad |||K_N * f|||_{\gamma_{N \in \mathbb{Z}}^q}^q \|_{\ell^p} \lesssim \sum_s |||L_{s,N} * f|||_{\gamma_{N \in \mathbb{Z}}^q}^q \|_{\ell^p} + |||(K_N - L_N) * f|||_{\ell_{N \in \mathbb{Z}}^1}^q \|_{\ell^p}.$$

The first sum in (4.9) is bounded by

$$\sum_s \mathbf{s}_s |\mathcal{R}_s|^{2|\frac{1}{p} - \frac{1}{2}| + \delta} \|f\|_{\ell^p} \lesssim \|f\|_{\ell^p}$$

by Proposition 4.2 and (4.7). Note that this estimate also holds for some  $1 < p_0 < \infty$  which is farther away from 2 than  $p$  and implies in particular

$$\|L_N\|_{\ell^{p_0} \rightarrow \ell^{p_0}} \leq C, \quad N \in \mathbb{Z}.$$

Since for our kernels also

$$\|K_N\|_{\ell^{p_0} \rightarrow \ell^{p_0}} \leq \|K_N\|_{\ell^1} \leq C$$

by (4.8), we obtain

$$\|K_N - L_N\|_{\ell^{p_0} \rightarrow \ell^{p_0}} \leq C.$$

Interpolating between  $\ell^{p_0}$  and  $\ell^2$  we obtain

$$\|K_N - L_N\|_{\ell^p \rightarrow \ell^p} \leq \|K_N - L_N\|_{\ell^{p_0} \rightarrow \ell^{p_0}}^{1-\theta} \|K_N - L_N\|_{\ell^2 \rightarrow \ell^2}^\theta \lesssim \|\widehat{K_N} - \widehat{L_N}\|_\infty^\theta,$$

where  $\theta > 0$  is obtained from the condition  $1/p = (1 - \theta)/p_0 + \theta/2$ . Interpolation is not needed if  $p = 2$ , and the condition (4.8) is consequently not used in that case. The last term is summable in  $N$  by (4.6). This, and Minkowski's inequality, allows us to estimate the second term in (4.9).  $\square$

**4.3. Short variation.** Since we will be able to handle the long variation on fairly dense subsets of  $\mathbb{N}$ , namely

$$(4.10) \quad Z = Z_\epsilon = \{ \lfloor 2^{k^\epsilon} \rfloor \}_{k=1}^\infty,$$

we can afford estimating the short variation in a very simplistic manner.

**Lemma 4.11.** *Let  $1 < p < \infty$ ,  $1 < q \leq \infty$ , let  $Z = \{N_1, N_2, \dots\} \subset \mathbb{N}$  be an increasing sequence, and suppose*

$$(4.12) \quad \left\| \sum_{N \in \mathcal{I}_k} \|K_{N+1} - K_N\|_{\ell^1} \right\|_{\ell_k^{\min(p,q)}} < \infty,$$

where  $\mathcal{I}_k = [N_k, N_{k+1}]$ . Then

$$|||K_N * f|||_{s\gamma_N^q}^q \|_{\ell^p} \lesssim_{p,q} \|f\|_{\ell^p}.$$

*Proof.* In view of the monotonicity of the variation norms it suffices to consider  $q \leq p$ . By the definition of the short variation norm, the monotonicity of variation norms, and two applications of the Minkowski inequality we have

$$(4.13) \quad \begin{aligned} |||K_N * f|||_{s\gamma_N^q}^q \|_{\ell^p} &= |||K_N * f|||_{\gamma_{N \in \mathcal{I}_j}^q}^q \|_{\ell_j^q}^q \|_{\ell^p} \leq |||K_N * f|||_{\gamma_{N \in \mathcal{I}_j}^1}^1 \|_{\ell_j^q}^q \|_{\ell^p} \\ &\leq |||(K_{N+1} - K_N) * f|||_{\ell_{N \in \mathcal{I}_j}^1}^1 \|_{\ell_j^q}^q \|_{\ell^p} \leq |||(K_{N+1} - K_N) * f|||_{\ell^p} \|_{\ell_{N \in \mathcal{I}_j}^1}^1 \|_{\ell_j^q}^q. \end{aligned}$$

By the Young convolution inequality this is bounded by

$$|||K_{N+1} - K_N|||_{\ell^1}^1 \|_{\ell_{N \in \mathcal{I}_j}^1}^1 \|_{\ell_j^q}^q \|f\|_{\ell^p}$$

By the hypothesis (4.12) this is  $\lesssim \|f\|_{\ell^p}$ .  $\square$

**Corollary 4.14.** *Let  $1 < p < \infty$  and suppose that for some  $Z \subset \mathbb{N}$  the conditions (4.6), (4.7), and (4.12) hold. If  $p \neq 2$  assume in addition (4.8). Then for every  $q > 2$  we have*

$$|||K_N * f|||_{\gamma_N^q}^q \|_{\ell^p} \lesssim_{p,q} \|f\|_{\ell^p}.$$

*Proof.* By Theorem 4.5 and Lemma 4.11 we have  $\ell^p$  bounds for  $\|K_N * f\|_{\sqrt[q]{N} \in \mathbb{Z}}$  and  $\|K_N * f\|_{\sqrt[q]{N}}$  with respect to  $Z$ . The conclusion follows from the pointwise bound (2.4).  $\square$

## 5. PRIMES

In this section we recall several estimates from [Bou88a, §4] and [Wie88], partially following the exposition in [MT13], and prove Theorem 1.1.

**5.1. Tools.** We begin with the necessary tools from number theory. In this section  $\Lambda$  denotes the von Mangoldt function,  $\mu$  the Möbius function, and  $\varphi$  the Euler totient function. For  $q \in \mathbb{N}$  let  $A_q = \{r \in \{1, \dots, q\} : (r, q) = 1\}$ . Recall the Ramanujan sum identity [Nat96, Theorem A.24]

$$(5.1) \quad \sum_{r \in A_q} e(ra/q) = \mu(q), \quad (a, q) = 1.$$

and the elementary estimate

$$(5.2) \quad \varphi(n) \gtrsim_\delta n^{1-\delta} \quad \text{for any } \delta > 0$$

for the Euler totient function, see [Nat96, Theorem A.16].

**Theorem 5.3** (Vinogradov, see [Dav80, §25]). *Suppose  $|\alpha - a/q| \leq 1/q^2$ ,  $a \in A_q$ . Then*

$$\left| \sum_{n \leq N} \Lambda(n) e(n\alpha) \right| \lesssim (Nq^{-1/2} + N^{4/5} + N^{1/2}q^{1/2})(\log N)^4.$$

**Theorem 5.4** (Siegel–Walfisz, see [Dav80, §22]). *Let*

$$\psi(N; q, r) = \sum_{n \leq N, n \equiv r \pmod{q}} \Lambda(n)$$

*be the (von Mangoldt weighted) counting function for the primes  $\equiv r \pmod{q}$ . For every  $A > 0$  there exists  $C(A) > 0$  such that*

$$\psi(N; q, r) = \frac{N}{\varphi(q)} + O(N \exp(-C(A)\sqrt{\log N}))$$

*for every  $q \leq (\log N)^A$  and  $r \in A_q$ .*

## 5.2. Approximation of the kernel.

**Lemma 5.5.** *Let  $A > 0$ ,  $0 < \epsilon < 1$ , and  $\alpha = a/q + \beta$ , where  $q \leq ((1 - \epsilon)\log N)^A$  and  $|\beta| < (\log N)^A/N$ . Then*

$$\left| \widehat{K_N}(\alpha) - \frac{\mu(q)}{\varphi(q)} \frac{1}{N} \sum_{n \leq N} e(n\beta) \right| \lesssim_A e^{-C(A)(1-\epsilon)\sqrt{\log N}},$$

*where  $C(A)$  is the constant from the Siegel–Walfisz theorem.*

The partial summation argument below is adapted from [MT13].

*Proof.* We write the von Mangoldt function as the increment of the weighted prime counting function

$$\Lambda(n) = \sum_{r=1}^q (\psi(n; q, r) - \psi(n-1; q, r)).$$

The terms with  $r \notin A_q$  are non-zero only for those  $n$  that are powers of the primes that divide  $q$ , and there are at most  $q \log N$  such  $n$ 's. Therefore

$$(5.6) \quad \widehat{K}_N(\alpha) = \frac{1}{N} \sum_{n \leq N} \Lambda(n) e(n\alpha) \\ = \frac{1}{N} \sum_{r \in A_q} e(ra/q) \sum_{n=M+1}^N (\psi(n; q, r) - \psi(n-1; q, r)) e(n\beta) + O\left(\frac{(\log N)^{4+2}}{N}\right) + O(M/N).$$

With  $M = \lceil N^{1-\epsilon} \rceil$  both error terms can be absorbed into the error term of the conclusion. Now consider the sum over  $n$  in the main term. By partial summation it equals

$$(5.7) \quad \psi(N; q, r) e(N\beta) - \psi(M; q, r) e(M\beta) + \sum_{n=M+1}^{N-1} \psi(n; q, r) (e(n\beta) - e((n+1)\beta)).$$

We use Theorem 5.4 to split this into a main term and the error term. The main term equals

$$\frac{N}{\varphi(q)} e(N\beta) - \frac{M}{\varphi(q)} e(M\beta) + \sum_{n=M+1}^{N-1} \frac{n}{\varphi(q)} (e(n\beta) - e((n+1)\beta)) = \frac{1}{\varphi(q)} \sum_{n=M}^N e(n\beta)$$

by partial summation. Summing up the contributions of these terms to (5.6) we obtain

$$\frac{1}{N} \sum_{r \in A_q} e(ra/q) \frac{1}{\varphi(q)} \sum_{n=M}^N e(n\beta) = \frac{\mu(q)}{\varphi(q)} \frac{1}{N} \sum_{n \leq N} e(n\beta) + O(M/N)$$

by the Ramanujan sum identity (5.1). It remains to estimate the error term produced by application of Theorem 5.4 to (5.7). It equals

$$O(Ne^{-C(A)\sqrt{\log N}}) + O(Me^{-C(A)\sqrt{\log M}}) + \sum_{n=M+1}^{N-1} O(ne^{-C(A)\sqrt{\log n}}) |e(n\beta) - e((n+1)\beta)| \\ = O(Ne^{-C(A)\sqrt{\log M}}) + |\beta| \sum_{n=M+1}^{N-1} O(Ne^{-C(A)\sqrt{\log M}}) \\ = O(Ne^{-C(A)\sqrt{\log M}}) + \frac{(\log N)^A}{N} O(N^2 e^{-C(A)\sqrt{\log M}}) \\ = O((\log N)^A N e^{-C(A)\sqrt{1-\epsilon}\sqrt{\log N}}).$$

The contribution of this term to (5.6) can therefore be estimated by

$$q(\log N)^A e^{-C(A)\sqrt{1-\epsilon}\sqrt{\log N}} \lesssim e^{-C(A)(1-\epsilon)\sqrt{\log N}}. \quad \square$$

**Lemma 5.8** ([Wie88, (6)]). *Let  $K_N$  be given by (1.2) and let  $L_N$  be given by (4.4) with  $d = 1$  and*

$$(5.9) \quad S(\theta) = \frac{\mu(q)}{\varphi(q)} \quad \text{for } \theta = a/q \text{ in reduced form.}$$

*Then for every  $B > 0$  we have*

$$\|\widehat{K}_N - \widehat{L}_N\|_\infty \lesssim_B (\log N)^{-B}.$$

*Proof.* Let  $A$  be sufficiently large ( $A = 3B + 8$  will do) and assume without loss of generality that  $N$  is sufficiently large depending on  $A$ . Let  $\alpha \in [0, 1]$ . By Dirichlet's approximation theorem there exists a reduced fraction  $a/q$  such that

$$(5.10) \quad \left| \alpha - \frac{a}{q} \right| \leq \frac{1}{qQ} \quad \text{with } q \leq Q = N(\log N)^{-A}.$$

Let  $s_0$  be such that  $a/q \in \mathcal{R}_{s_0}$  and let  $\beta = \alpha - a/q$ . For each  $s \neq s_0$  there is at most one  $a_s/q_s \in \mathcal{R}_s$  that contributes to the sum defining  $\widehat{L_{s,N}}(\alpha)$ , and

$$(5.11) \quad \sum_{s \neq s_0} |\widehat{L_{s,N}}(\alpha)| \leq \sum_{s \neq s_0} \mathbf{S}_s \widehat{\sigma_N}(\alpha - a_s/q_s) \lesssim \sum_{s \neq s_0} \mathbf{S}_s (1 + N|\alpha - a_s/q_s|)^{-1}.$$

We have

$$(5.12) \quad \left| \alpha - \frac{a_s}{q_s} \right| \geq \left| \frac{a_s}{q_s} - \frac{a}{q} \right| - |\beta| \geq \frac{1}{qq_s} - \frac{1}{qQ}.$$

The first term dominates if  $2^s < Q/4$  and in particular if  $2^s \lesssim_A (\log N)^{A/2}$ . Thus we may estimate (5.11) by

$$\sum_{s: 2^s \lesssim_A (\log N)^{A/2}} N^{-1} q \cdot 2^s + \sum_{s: 2^s \gtrsim_A (\log N)^{A/2}} \mathbf{S}_s.$$

Using (5.10) in the first term and (5.2) in the second term we see that this is  $\lesssim (\log N)^{-A/2+\delta}$ . Thus it remains to show

$$|\widehat{L_{s_0,N}}(\alpha) - \widehat{K_N}(\alpha)| \lesssim_B (\log N)^{-B}.$$

Major arcs. Suppose that  $q \leq (\frac{1}{2} \log N)^A$ . The estimate (5.12) shows that  $|\alpha - \theta| > 4^{-s_0}/2$  for all  $\theta \in \mathcal{R}_{s_0} \setminus \{a/q\}$  (provided that  $N$  is large enough), so the corresponding terms in  $\widehat{L_{s_0,N}}$  vanish at  $\alpha$ , and we obtain

$$\widehat{L_{s_0,N}}(\alpha) = S(a/q) \widehat{\sigma_N}(\beta) \widehat{\eta_{10^{s_0}}}(\beta).$$

Further,

$$|\widehat{\eta_{10^{s_0}}}(\beta) - 1| \lesssim 10^{s_0} |\beta| \lesssim q^4 \frac{1}{qQ} \lesssim \frac{(\log N)^{4A}}{N}.$$

Finally,

$$\left| \widehat{\sigma_N}(\beta) - \frac{1}{N} \sum_{n \leq N} e(n\beta) \right| \lesssim |\beta| \leq \frac{(\log N)^A}{N}.$$

Combining these estimates with Lemma 5.5 we obtain the desired bound at  $\alpha$ .

Minor arcs. Suppose now  $q > (\frac{1}{2} \log N)^A$ . Then

$$|\widehat{L_{s_0,N}}(\alpha)| \leq \mathbf{S}_{s_0} \lesssim q^{-1+\delta} \lesssim (\log N)^{(-1+\delta)A}.$$

On the other hand, by Theorem 5.3 we have

$$\begin{aligned} |\widehat{K_N}(\alpha)| &\lesssim (q^{-1/2} + N^{-1/5} + N^{-1/2} q^{1/2}) (\log N)^4 \\ &\lesssim ((\log N)^{-A/2} + N^{-1/5}) (\log N)^4 \lesssim (\log N)^{-A/2+4}. \quad \square \end{aligned}$$

*Proof of Theorem 1.1 for  $|\frac{1}{p} - \frac{1}{2}| < \frac{1}{4}$ .* It suffices to verify the conditions of Corollary 4.14 with  $d = 1$ . We consider  $Z = Z_\epsilon$  as in (4.10) with a sufficiently small  $\epsilon$  to be chosen shortly. We have

$$\|K_{N+1} - K_N\|_{\ell^1} \leq \frac{1 + \log(N+1)}{N+1}.$$

For  $\mathcal{I}_k = [N_k, N_{k+1}]$  we also have

$$|\mathcal{I}_k| \lesssim 2^{k^\epsilon} (2^{(k+1)^\epsilon - k^\epsilon} - 1) \lesssim 2^{k^\epsilon} k^{-1+\epsilon}$$

This gives the bound

$$\|2^{k^\epsilon} k^{-1+\epsilon} (1 + \log((2^{(k+1)^\epsilon} + 1)))/2^{k^\epsilon}\|_{\ell_k^{\min(p,q)}} \lesssim \|k^{-1+2\epsilon}\|_{\ell_k^{\min(p,q)}}$$

for (4.12). This is finite provided  $(1 - 2\epsilon) \min(p, q) > 1$ .

The condition (4.8) is immediate.

With the choice (5.9) we have

$$\mathbf{s}_s \lesssim 2^{-s(1-\delta)}$$

for any  $\delta > 0$  by (5.2). Since  $|\mathcal{R}_s| \leq 4^s$ , this gives (4.7) whenever  $\left|\frac{1}{p} - \frac{1}{2}\right| < \frac{1}{4}$ . Finally, (4.6) follows from Lemma 5.8.  $\square$

**5.3. A multi-frequency estimate for  $1 < p < \infty$ .** In view of the results in [NOT10] it appears plausible that the exponent  $2\left|\frac{1}{p} - \frac{1}{2}\right| + \delta$  in Proposition 3.9 can be improved to  $\left|\frac{1}{p} - \frac{1}{2}\right| + \delta$ . If this is indeed the case, then the above proof immediately gives Theorem 1.1 for  $1 < p < \infty$ .

As an aside we note that the required improvement of Proposition 3.9 can be obtained for  $p > 2$  by proving a version of Corollary 3.6 for Hilbert space-valued functions, which seems to be possible using the methods in [JSW08], and applying it together with Rubio de Francia's Littlewood–Paley inequality for arbitrary intervals [RdF85] to obtain the necessary endpoint estimates for  $p$  near  $\infty$ .

However, this approach does not extend to  $p < 2$ , and, in any case, a much simpler argument due to Wierdl [Wie88] works for our purposes. The main additional ingredient is the Magyar–Stein–Wainger result on periodic multipliers (Theorem 4.1), which can be used to close a gap on p. 331 in [Wie88]: there, the proof of the estimate for  $(**)$  gives  $q^p$  instead of  $q$ . (That gap has been already closed in [MT13] using what amounts to a special case of Theorem 4.1. We note that this gap does not affect the validity of the results in [Wie88] since the discrepancy between  $q$  and  $q^p$  can be absorbed into the estimates that follow for  $p$  near 1 and interpolation with the easy endpoint at  $p = \infty$  allows one to handle large values of  $p$ . However, the situation for variational estimates is different due to lack of such an easy endpoint, and the full power of Theorem 4.1 is useful here.)

**Lemma 5.13.** *For  $q \in \mathbb{N}$  and  $Q \in \mathbb{R}_{>0}$  let*

$$\widehat{L_{q,Q,t}}(\vec{a}) := \sum_{a_1, \dots, a_d=1}^q \widehat{\sigma}_t(\vec{a} - \frac{a}{q}) \widehat{\eta}_Q(\vec{a} - \frac{a}{q}).$$

*Suppose  $q \leq 25Q$ ,  $r > 2$ ,  $1 < p < \infty$ . Then*

$$\| \|L_{q,Q,t} * f\|_{\mathcal{V}_{t>0}^r} \|_{\ell^p(\mathbb{Z}^d)} \lesssim_{p,r} \|f\|_{\ell^p(\mathbb{Z}^d)}.$$

*Proof.* As usually, it suffices to consider a finite set  $T$  of  $t$ 's as long as the bound does not depend on  $T$ . The estimate follows from the single-frequency estimate in Corollary 3.6 using Theorem 4.1 with  $B_1 = \mathbb{C}$ ,  $B_2 = (\mathbb{C}^T, \mathcal{V}_{t \in T}^r)$ ,  $q = q$ , and  $p = p$ .  $\square$

**Corollary 5.14.** *Let*

$$\widehat{L_{A_q^d, Q, t}}(\vec{a}) := \sum_{a \in A_q^d} \widehat{\sigma}_t(\vec{a} - \frac{a}{q}) \widehat{\eta}_Q(\vec{a} - \frac{a}{q}),$$

*where  $A_q^d = \{(a_1, \dots, a_d) : (a_1, \dots, a_d, q) = 1\}$ . Suppose  $q \leq 25Q$ ,  $q \in \mathbb{N}$ ,  $Q \in \mathbb{R}_{>0}$ ,  $r > 2$ ,  $1 < p < \infty$ . Then*

$$\| \|L_{q,Q,t} * f\|_{\mathcal{V}_{t>0}^r} \|_{\ell^p(\mathbb{Z}^d)} \lesssim_{p,r,\epsilon} q^\epsilon \|f\|_{\ell^p(\mathbb{Z}^d)}$$

*for any  $\epsilon > 0$ .*

*Proof.* Recall [HW08, Theorem 266] the Möbius inversion formula

$$g(n) = \sum_{d|n} f(d) \implies f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) g(d).$$

It is applied with  $f(n) = L_{A_n^d, Q, t}$ , then  $g(n) = L_{n, Q, t}$ , so

$$L_{A_q^d, Q, t} = \sum_{d|q} \mu\left(\frac{q}{d}\right) L_{d, Q, t}.$$

For each summand we have the estimate  $\lesssim \|f\|_{\ell^p} \leq \|f\|_{\ell^p}$  by Lemma 5.13 and there are  $O(q^\epsilon)$  summands since  $q$  has  $O(q^\epsilon)$  divisors for any  $\epsilon > 0$  [HW08, Theorem 315].  $\square$

**Corollary 5.15.** *Let  $d = 1$  and  $S(a/q) = \mu(q)/\varphi(q)$ . Suppose  $r > 2$ ,  $1 < p < \infty$ . Then*

$$\| \| L_{s, t} * f \|_{\gamma_{t \in \mathbb{R}}^r} \|_{\ell^p} \lesssim_{p, r, \epsilon} 2^{s\epsilon} \|f\|_{\ell^p}$$

for any  $\epsilon > 0$ .

*Proof.* We have

$$L_{s, t} = \sum_{q=2^s}^{2^{s+1}-1} \frac{\mu(q)}{\varphi(q)} L_{A_q, 10^s, t},$$

so

$$\| \| L_{s, t} * f \|_{\gamma_{t \in \mathbb{R}}^r} \|_{\ell^p} \leq \sum_{q=2^s}^{2^{s+1}-1} \frac{1}{\varphi(q)} \| \| L_{A_q, 10^s, t} * f \|_{\gamma_{t \in \mathbb{R}}^r} \|_{\ell^p}.$$

By Corollary 5.14 with  $d = 1$  and (5.2) this is bounded by

$$\sum_{q=2^s}^{2^{s+1}-1} \frac{1}{q^{1-\epsilon}} q^\epsilon \|f\|_{\ell^p} \lesssim (2^s)^{2\epsilon} \|f\|_{\ell^p}$$

for any  $\epsilon > 0$ .  $\square$

*Proof of Theorem 1.1 for  $1 < p < \infty$ .* An inspection of §4 reveals that it suffices to prove a bound for

$$\| \| L_{s, t} * f \|_{\gamma_{t \in \mathbb{R}}^r} \|_{\ell^p}$$

that is summable in  $s$ . This is given by interpolation between Corollary 5.15 and Proposition 4.2.  $\square$

## 6. POLYNOMIALS

In order to make the presentation self-contained we summarize here the approximation of the kernel (1.4) in terms of objects introduced in §4 following the argument in [Bou89]. We then prove Theorem 1.3.

We denote by  $\delta > 0$  a small quantity that depends only on  $d$ . This symbol's value may change between its uses. Also, the implied constant in the  $\lesssim$  notation may depend on  $d$ ; all other dependencies will be noted explicitly.

**6.1. Estimates for trigonometric sums.** Both the minor and the major arc estimates in this section rely on the following estimate for complete exponential sums, which is due to Hua.

**Lemma 6.1** ([Hua40]). *Let  $\theta_j = a_j/q_j$ ,  $\text{lcm}(q_1, \dots, q_d) = N$ . Then for any  $\delta > 0$  we have*

$$|\widehat{K_N}(\vec{\theta})| \lesssim_{d, \delta} N^{-1/d+\delta}.$$

The major arcs used in the estimate for  $\widehat{K_N}$  are

(6.2)

$$\mathfrak{M}_N = \{\vec{\alpha} \in \mathbb{R}^d : \alpha_j = a_j/q_j + \beta_j, (a_j, q_j) = 1, |\beta_j| \leq N^{-j+\nu}, 1 \leq j \leq d, \text{lcm}(q_1, \dots, q_d) \leq N^\nu\},$$

where  $\nu = 1/\max(d, 12)$ . On the minor arcs we have the following estimate.



**Lemma 6.3** ([Vin71, Chapter IV, Theorem 3 (p. 72)]). *There exists  $\delta = \delta(d) > 0$  such that for any  $N$  and any  $\vec{\alpha} \notin \mathfrak{M}_N$  we have*

$$|\widehat{K_N}(\vec{\alpha})| \lesssim N^{-\delta}.$$

In Vinogradov's 1971 book (which is almost, but not quite, entirely unlike the homonymous 1947 publication and its 1954 English translation) this result is stated for  $d \geq 12$ , which clearly implies the cases  $d < 12$ .

## 6.2. Approximation of trigonometric sums.

**Lemma 6.4** ([Bou89, Lemma 5.12]). *Let  $\vec{\alpha} \in \mathfrak{M}_N$ . Then*

$$\widehat{K_N}(\vec{\alpha}) = \widehat{K_q}(\vec{\theta}) \widehat{\sigma_N}(\vec{\beta}) + O(N^{-\delta}) \quad \text{for some } \delta > 0,$$

where  $q = \text{lcm}(q_1, \dots, q_d)$  and  $\vec{\beta}$  is as in (6.2).

*Proof.* For  $n = qs + r$ ,  $1 \leq s \leq \lfloor N/q \rfloor$ ,  $1 \leq r \leq q$ , it follows from the assumptions that

$$\alpha_j n^j = \theta_j r^j + \beta_j s^j q^j + O(N^{-\delta}).$$

Hence the average in the definition of  $\widehat{K_N}(\vec{\alpha})$  can be approximated by a product of averages over  $s$  and  $r$ , namely

$$\widehat{K_N}(\vec{\alpha}) = \widehat{K_q}(\vec{\theta}) \widehat{K_{\lfloor N/q \rfloor}}(\beta_1 q^1, \dots, \beta_d q^d) + O(N^{-\delta}).$$

Now,  $\widehat{K_{\lfloor N/q \rfloor}}(\beta_1 q^1, \dots, \beta_d q^d)$  is a Riemann sum for  $\widehat{\sigma_N}(\vec{\beta})$  at scale  $q$ . The assumptions on  $|\beta_j|$  imply that the derivative of the integrand in  $\widehat{\sigma_N}(\vec{\beta})$  is  $O(N^{-\delta})$ , and the boundary effects are of the same order, so the error incurred by passing from the sum to the integral is  $O(N^{-\delta})$ .  $\square$

**Lemma 6.5** ([Bou89, Lemma 5.17]). *For every  $1 \leq j \leq d$  we have*

$$(6.6) \quad |\widehat{\sigma_N}(\vec{\beta})| \lesssim_{j,d} |\beta_j|^{-1/d} N^{-j/d}.$$

*Proof.* Recall

$$\widehat{\sigma_N}(\vec{\beta}) = \frac{1}{N} \int_{s=0}^N e(\phi(x)) dx,$$

where  $\phi(x) = \beta_1 x^1 + \dots + \beta_d x^d$ . We will show (6.6) by descending induction on  $j$ . Suppose that (6.6) is known for all  $j > k$ , we have to show (6.6) with  $j = k$ . If  $|\beta_j|^{-1/d} N^{-j/d} \lesssim |\beta_k|^{-1/d} N^{-k/d}$  for some  $j > k$ , then this follows from the induction hypothesis. Otherwise we have  $|\beta_j| N^j \lesssim |\beta_k| N^k$ , so that

$$|\phi^{(k)}(x)| = |k! \beta_k + \sum_{j=k+1}^d \frac{j!}{(j-k)!} \beta_j x^{j-k}| > |k! \beta_k| - \sum_{j=k+1}^d \frac{j!}{(j-k)!} |\beta_j| N^{j-k} \gtrsim |\beta_k|$$

for all  $x \in [0, N]$  provided that the implied constant in the  $\lesssim$  notation was chosen sufficiently small. The van der Corput estimate [Ste93, §VIII.1.2, Proposition 2] then implies the desired conclusion (note that the van der Corput estimate is applicable also for  $k = 1$  since  $\phi''$  changes sign at most  $d$  times).  $\square$

**Proposition 6.7** (cf. [Bou89, Lemma 6.14]). *Let  $L_N$  be given by (4.4) with*

$$(6.8) \quad S(\vec{\theta}) = \widehat{K_q}(\vec{\theta}), \quad \text{where } q \text{ is the least common denominator of } \vec{\theta}.$$

and let  $K_N$  be given by (1.4). Then

$$\|\widehat{K_N} - \widehat{L_N}\|_{\infty} \lesssim N^{-\delta} \quad \text{for some } \delta > 0.$$

*Proof.* We may assume  $N > 10000$ , say. Suppose first  $\vec{\alpha} \in \mathfrak{M}_N$ . Let  $\theta_j = a_j/q_j$  and  $\beta_j = \alpha_j - \theta_j$  be as in (6.2). By Lemma 6.4 we have

$$\widehat{K}_N(\vec{\alpha}) = S(\vec{\theta})\widehat{\sigma}_N(\vec{\beta}) + O(N^{-\delta}).$$

Let  $s_0$  be such that  $\vec{\theta} \in \mathcal{R}_{s_0}$  and let  $s_1$  be the integer such that  $N^\nu \leq 2^{s_1} < 2N^\nu$ . Then

$$|S(\vec{\theta})\widehat{\sigma}_N(\vec{\beta}) - \sum_{s \geq 0} \widehat{L}_{s,N}(\vec{\alpha})| \lesssim |1 - \widehat{\eta}_{10^{s_0}}(\vec{\beta})| + \sum_{0 \leq s \leq s_1} \sup_{\vec{\theta}' \in \mathcal{R}_s, \vec{\theta}' \neq \vec{\theta}} |\widehat{\sigma}_N(\vec{\alpha} - \vec{\theta}')| + \sum_{s > s_1} \sup_{\vec{\theta}' \in \mathcal{R}_s} |S(\vec{\theta}')|.$$

Since  $|\beta_j| \leq N^{-j+\nu} \leq N^{-4\nu}/100 \leq 10^{-s_0}/100$ , the first term vanishes. We estimate the third term using Lemma 6.1 and the second term using (6.6) and the observation that for  $\vec{\theta}' \in \mathcal{R}_s$ ,  $s \leq s_1$ , we have

$$|\alpha_j - \theta'_j| \geq |\theta'_j - \theta_j| - |\alpha_j - \theta_j| \gtrsim 2^{-s_1-s_0} - N^{-j+\nu} \gtrsim N^{-2\nu}$$

for any  $j$  with  $\theta'_j \neq \theta_j$ . This yields the estimate

$$|\widehat{K}_N(\vec{\alpha}) - \sum_{s \geq 0} \widehat{L}_{s,N}(\vec{\alpha})| \lesssim N^{-\delta} + \sum_{0 \leq s \leq s_1} (N^{-2\nu} \cdot N)^{-1/d} + \sum_{s > s_1} (2^s)^{-\delta} \lesssim N^{-\delta}$$

as required.

If  $\vec{\alpha} \notin \mathfrak{M}_N$ , then we have  $|\widehat{K}_N(\vec{\alpha})| \lesssim N^{-\delta}$  by Lemma 6.3. Let  $s_1$  be the integer such that  $N^\nu/2 \leq 2^{s_1} < N^\nu$ . We have

$$|\sum_{s \geq 0} \widehat{L}_{s,N}(\vec{\alpha})| \lesssim \sum_{0 \leq s < s_1} \sup_{\vec{\theta} \in \mathcal{R}_s} |\widehat{\sigma}_N(\vec{\alpha} - \vec{\theta})| + \sum_{s \geq s_1} \sup_{\vec{\theta} \in \mathcal{R}_s} |S(\vec{\theta})|.$$

The second summand is  $\lesssim N^{-\delta}$  as before. In the first summand note that  $|\alpha_j - \theta_j| > N^{-j+\nu}$  for some  $j$  since otherwise we would have  $\vec{\alpha} \in \mathfrak{M}_N$ . In view of (6.6) this implies

$$|\widehat{\sigma}_N(\vec{\alpha} - \vec{\theta})| \lesssim N^{-\nu/d},$$

and, since the first summand above consists of approximately  $\log N$  terms, we are done.  $\square$

*Proof of Theorem 1.3 for  $\left|\frac{1}{p} - \frac{1}{2}\right| < \frac{1}{4d^2}$ .* It suffices to verify the conditions of Corollary 4.14 with  $Z = Z_\epsilon$  as in (4.10) for a sufficiently small  $\epsilon$ . Verification of (4.12) is similar to, but easier than, the proof of Theorem 1.1, and the condition (4.8) is trivially satisfied.

By Lemma 6.1 we have

$$\mathbf{S}_s \lesssim_\delta 2^{s(-1/d+\delta)} \quad \text{for any } \delta > 0,$$

and since  $|\mathcal{R}_s| \leq 4^{(s+1)d}$  this implies (4.7) provided  $\left|\frac{1}{p} - \frac{1}{2}\right| < \frac{1}{4d^2}$ . Finally, condition (4.6) is given by Proposition 6.7.  $\square$

**6.3. Another multi-frequency estimate for  $1 < p < \infty$ .** The difficulty in the case of polynomials stems from the fact that the estimates for the exponential sums  $S(\vec{a}/q)$  are much worse than in the case of the primes. We have the following version of Lemma 5.13

**Lemma 6.9.** *For  $q \in \mathbb{N}$  and  $Q \in \mathbb{R}_{>0}$  let*

$$\widehat{L}_{q,Q,t,\text{poly}}(\vec{\alpha}) := \sum_{a_1, \dots, a_d=1}^q S(a/q) \widehat{\sigma}_t(\vec{\alpha} - \frac{a}{q}) \widehat{\eta}_Q(\vec{\alpha} - \frac{a}{q}),$$

where the coefficients  $S$  are given by (6.8). Suppose  $q \leq 5Q$ ,  $r > 2$ ,  $1 < p < \infty$ . Then

$$\| \| L_{q,Q,t,\text{poly}} * f \|_{\mathcal{V}_{t \in \mathbb{R}}^r} \|_{\ell^p(\mathbb{Z}^d)} \lesssim_{p,r} \| f \|_{\ell^p(\mathbb{Z}^d)}.$$

*Proof.* With the notation from Lemma 5.13 we have

$$\widehat{L_{q,Q,t,\text{poly}}}(\vec{a}) = \widehat{L_{q,Q,t}}(\vec{a}) \left( \sum_{a_1, \dots, a_d=1}^q S(a/q) \widehat{\eta_{Q/4}}(\vec{a} - \frac{a}{q}) \right).$$

In view of that lemma it suffices to show that the inverse Fourier transform of the function in the brackets is uniformly bounded in  $\ell^1(\mathbb{Z}^d)$ . The value of that inverse Fourier transform at  $x \in \mathbb{Z}^d$  equals

$$\begin{aligned} & \int_{(\mathbb{R}/\mathbb{Z})^d} \sum_{a_1, \dots, a_d=1}^q S(a/q) \widehat{\eta_{Q/4}}(\vec{a} - \frac{a}{q}) e(\alpha_1 x_1 + \dots + \alpha_d x_d) d\alpha_1 \dots d\alpha_d \\ &= \sum_{a_1, \dots, a_d=1}^q S(a/q) \int_{(\mathbb{R}/\mathbb{Z})^d} \widehat{\eta_{Q/4}}(\vec{a}) e(\alpha_1 x_1 + \dots + \alpha_d x_d) e(\frac{a_1}{q} x_1 + \dots + \frac{a_d}{q} x_d) d\alpha_1 \dots d\alpha_d \\ &= \eta_{Q/4}(x) \sum_{a_1, \dots, a_d=1}^q S(a/q) e(\frac{a_1}{q} x_1 + \dots + \frac{a_d}{q} x_d) \\ &= \eta_{Q/4}(x) \sum_{a_1, \dots, a_d=1}^q \frac{1}{q} \sum_{n=1}^q e(-\frac{a_1}{q} n^1 - \dots - \frac{a_d}{q} n^d) e(\frac{a_1}{q} x_1 + \dots + \frac{a_d}{q} x_d) \\ &= \eta_{Q/4}(x) \frac{1}{q} \sum_{n=1}^q \prod_{j=1}^d \sum_{a_j=1}^q e(\frac{a_j}{q} (x_j - n^j)) \\ &= \eta_{Q/4}(x) q^{d-1} \sum_{n=1}^q \prod_{j=1}^d 1_{x_j \equiv n^j \pmod{q}} \\ &= \eta_{Q/4}(x) q^{d-1} 1_{x_j \equiv x_1^j \pmod{q}, j \geq 2}. \end{aligned}$$

This is uniformly bounded in  $\ell^1$  for  $q \lesssim Q$ . To see this note that the characteristic function selects exactly  $q$  points from every cube with side length  $q$ , so we have

$$\begin{aligned} \|\eta_{Q/4}(x) q^{d-1} 1_{x_j \equiv x_1^j \pmod{q}, j \geq 2}\|_{\ell^1} &\leq q^d \sum_{y \in \mathbb{Z}^d} \sup_{x \in qy + [1, q]^d} \eta_{Q/4}(x) \\ &\lesssim q^d Q^{-d} \sum_{y \in \mathbb{Z}^d} \sup_{x \in qy + [1, q]^d} \eta(4x/Q) \\ &\lesssim q^d Q^{-d} \sum_{y \in \mathbb{Z}^d} \sup_{x \in qy + [1, q]^d} (1 + |x|/Q)^{-d-1} \\ &\lesssim q^d Q^{-d} \sum_{y \in \mathbb{Z}^d, |y| < CQ/q} 1 + q^d Q^{-d} \sum_{y \in \mathbb{Z}^d, |y| \geq CQ/q} (q|y|/Q)^{-d-1} \\ &\lesssim 1. \end{aligned} \quad \square$$

*Proof of Theorem 1.3 for  $\left| \frac{1}{p} - \frac{1}{2} \right| < \frac{1}{2(d+1)}$ .* Repeating the argument leading to Corollary 5.15 we obtain

$$\|L_{s,t} * f\|_{\mathcal{V}_{t>0}^r} \|\cdot\|_{\ell^p(\mathbb{Z}^d)} \lesssim_{p,r,\epsilon} 2^{s(1+\epsilon)} \|f\|_{\ell^p(\mathbb{Z}^d)}$$

for  $1 < p < \infty$ . Recall that in view of Proposition 4.2 we can replace  $(1 + \epsilon)$  by  $(-1/d + \epsilon)$  for  $p = 2$ . Interpolation gives

$$\|L_{s,t} * f\|_{\mathcal{V}_{t>0}^r} \|\cdot\|_{\ell^p(\mathbb{Z}^d)} \lesssim_{p,r,\epsilon} 2^{s(-(1-|1-2/p|)/d + |1-2/p| + \epsilon)} \|f\|_{\ell^p(\mathbb{Z}^d)},$$

and this is summable provided

$$|1/2 - 1/p| < 1/2(d+1).$$

The remaining ingredients of the proof apply unchanged. □

## REFERENCES

- [Bou88a] J. Bourgain. “An approach to pointwise ergodic theorems”. In: *Geometric aspects of functional analysis (1986/87)*. Vol. 1317. Lecture Notes in Math. Berlin: Springer, 1988, pp. 204–223.
- [Bou88b] J. Bourgain. “On the maximal ergodic theorem for certain subsets of the integers”. In: *Israel J. Math.* 61.1 (1988), pp. 39–72.
- [Bou88c] J. Bourgain. “On the pointwise ergodic theorem on  $L^p$  for arithmetic sets”. In: *Israel J. Math.* 61.1 (1988), pp. 73–84.
- [Bou89] J. Bourgain. “Pointwise ergodic theorems for arithmetic sets”. In: *Inst. Hautes Études Sci. Publ. Math.* 69 (1989). With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein, pp. 5–45.
- [Cal68] A.-P. Calderón. “Ergodic theory and translation-invariant operators”. In: *Proc. Nat. Acad. Sci. U.S.A.* 59 (1968), pp. 349–353.
- [Dav80] H. Davenport. “Multiplicative number theory”. Second. Vol. 74. Graduate Texts in Mathematics. Revised by Hugh L. Montgomery. New York: Springer-Verlag, 1980, pp. xiii+177.
- [Hua40] L.-k. Hua. “On an exponential sum”. In: *J. Chinese Math. Soc.* 2 (1940), pp. 301–312.
- [HW08] G. H. Hardy and E. M. Wright. “An introduction to the theory of numbers”. Sixth. Revised by D. R. Heath-Brown and J. H. Silverman, With a foreword by Andrew Wiles. Oxford University Press, Oxford, 2008, pp. xxii+621.
- [JKRW98] R. L. Jones, R. Kaufman, J. M. Rosenblatt, and M. Wierdl. “Oscillation in ergodic theory”. In: *Ergodic Theory Dynam. Systems* 18.4 (1998), pp. 889–935.
- [JSW08] R. L. Jones, A. Seeger, and J. Wright. “Strong variational and jump inequalities in harmonic analysis”. In: *Trans. Amer. Math. Soc.* 360.12 (2008), pp. 6711–6742.
- [MSW02] A. Magyar, E. M. Stein, and S. Wainger. “Discrete analogues in harmonic analysis: spherical averages”. In: *Ann. of Math. (2)* 155.1 (2002), pp. 189–208. arXiv:math/0409365 [math.CA].
- [MT13] M. Mirek and B. Trojan. “Cotlar’s ergodic theorem along the prime numbers”. In: *J. Fourier Anal. Appl.* (2013). To appear. arXiv:1311.7572 [math.DS].
- [MT14] M. Mirek and B. Trojan. “Discrete maximal functions in higher dimensions and applications to ergodic theory”. Preprint. 2014. arXiv:1405.5566 [math.CA].
- [Nat96] M. B. Nathanson. “Additive number theory”. Vol. 164. Graduate Texts in Mathematics. The classical bases. New York: Springer-Verlag, 1996, pp. xiv+342.
- [NOT10] F. Nazarov, R. Oberlin, and C. Thiele. “A Calderón-Zygmund decomposition for multiple frequencies and an application”. In: *Math. Res. Lett.* 17.3 (2010), pp. 529–545. arXiv:0912.3010 [math.CA].
- [Obe13] R. Oberlin. “Bounds on the Walsh model for  $M^{q,*}$  Carleson and related operators”. In: *Rev. Mat. Iberoam.* 29.3 (2013), pp. 829–857.
- [RdF85] J. L. Rubio de Francia. “A Littlewood-Paley inequality for arbitrary intervals”. In: *Rev. Mat. Iberoamericana* 1.2 (1985), pp. 1–14.
- [Ste93] E. M. Stein. “Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals”. Vol. 43. Princeton Mathematical Series. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III. Princeton, NJ: Princeton University Press, 1993, pp. xiv+695.
- [Vin71] I. M. Vinogradov. “Metod trigonometricheskikh summ v teorii chisel”. Izdat. “Nauka”, Moscow, 1971, p. 159.
- [Wie88] M. Wierdl. “Pointwise ergodic theorem along the prime numbers”. In: *Israel J. Math.* 64.3 (1988), 315–336 (1989).
- [Kra14a] B. Krause. “Polynomial Ergodic Averages Converge Rapidly: Variations on a Theorem of Bourgain”. Preprint. 2014. arXiv:1402.1803 [math.CA].
- [Kra14b] B. Krause. “Some Optimizations for (Maximal) Multipliers in  $L^p$ ”. Preprint. 2014. arXiv:1402.1804 [math.CA].

INSTITUTE OF MATHEMATICS, HEBREW UNIVERSITY, GIVAT RAM, JERUSALEM, 91904, ISRAEL